AD-A095 869 SOUTH CAROLINA UNIV COLUMBIA F/G 12/1
ESTIMATION OF LOCATION DIFFERENCE FOR FRAGMENTARY SAMPLES.(U)
DEC 80 L J WEI F49620-79-C-0140
UNCLASSIFIED AFOSR-TR-81-0125 NL

1 OF 1
AD
A095869

END
DATE
FILMED
3 81
DTIC

AFOSR-TR-81-0125

ESTIMATION OF LOCATION DIFFERENCE FOR FRAGMENTARY SAMPLES

L. J. Wei
National Cancer Institute
and
University of South Carolina
Columbia, S.C. 29208

F49620-79-C-0140

SUMMARY

A class of simple and robust estimators of the difference between location parameters of correlated variables is proposed when some observations on either of the variables are missing. We show that these estimators are consistent, asymptotically normally distributed, and insensitive to outlying observations. Asymptotic relative efficiency comparisons with other known estimators are made to show the advantage of the proposed estimators.

Some keywords: Asymptotic relative efficiency; bivariate exponential distribution; consistent; Hodges-Lehmann estimator; median.

2

1. INTRODUCTION

The problem of estimating the difference between means of a bivariate normal distribution when some observations on either of the variables are missing has received a great deal of attention in recent statistical literature (Wilks, 1932; Anderson, 1957; Hocking and Smith, 1968; Mehta and Gurland, 1969; Lin, 1971, 1973; Lin and Stivers, 1974). In this article we study the problem of estimation of the difference between the location parameters of correlated variables from fragmentary samples when the population being sampled is not necessarily normal. More specifically, let $(X, Y-\theta)'$ be a random vector with absolutely continuous joint distribution function H which is free of θ and is symmetric in its arguments, i.e. $H(u,v) = H(v,u)$, for $(u,v)' \varepsilon R^2$. Also, let $(X_i, Y_i)'$, $i=1,\ldots,n$ be n pairs of observations on $(X,Y)'$; X_{n+j}, $j=1,\ldots,s$, be s additional observations on X; Y_{n+k}, $k=1,\ldots,t$, be t additional observations on Y. The $(X_i, Y_i)'$, X_{n+j}, and Y_{n+k} are assumed to be mutually independent for $i=1,\ldots,n$, $j=1,\ldots,s$ and $k=1,\ldots,t$. The problem is how to use the fragmentary sample in the most efficient way to estimate the shift parameter θ. Gupta and Rohatgi (1981) considered the case that X and Y are linearly related and constructed regression estimators which are linear combinations of fragmentary sample means. Therefore, their estimators are sensitive to outlying observations.

A class of simple and robust estimators $\hat{\theta}$ of θ is proposed in Section 2. We show that these estimators are unbiased if the underlying distribution H is symmetric about some point $(\mu_1, \mu_2)'$ or the two fragmentary sample sizes are equal, i.e., $s=t$. Also, it is shown that $\hat{\theta}$ is consistent and asymptotically normally distributed. In Section 3 we compare the asymptotic relative efficiency of $\hat{\theta}$ with other known estimators.

2. THE ESTIMATOR $\hat{\theta}$

For the fragmentary sample given in Section 1, we take all possible differences $Y_j - X_i$, $i=1,\ldots,n+s, j=1,\ldots,n+t$, each with multiplicity α. We also take all possible Walsh differences $\frac{1}{2}(Y_i+Y_{i'}-X_i-X_{i'})$, $1 \le i \le i' \le n$, each with multiplicity β. Denote the ordered set of the above $M=\alpha(n+s)(n+t)+\beta n(n+1)/2$ differences by $D_{(1)} \le \cdots \le D_{(M)}$ where α and β are nonnegative integers. A natural estimator $\hat{\theta}$ of θ is then the median of the D's, i.e.,

$$\hat{\theta} = \hat{\theta}(\alpha,\beta) = \begin{cases} \frac{1}{2}(D_{(\ell)}+D_{(\ell+1)}), & \alpha(n+s)(n+t)+\beta n(n+1)/2=2\ell, \\ D_{(\ell+1)}, & \alpha(n+s)(n+t)+\beta n(n+1)/2=2\ell+1. \end{cases}$$

Note that the estimator $\hat{\theta}(0,1)$ which disregards the information from the incomplete pairs is a Hodges-Lehmann estimator of θ based on Wilcoxon signed rank statistic. The estimator $\hat{\theta}(1,0)$ which uses all the data points but ignores the pairing information is based on the two-sample Mann-Whitney-Wilcoxon statistic.

First, it is proved in the Appendix that the distribution of the difference $\hat{\theta}(\alpha,\beta)-\hat{\theta}$ is free of θ and the estimator $\hat{\theta}(\alpha,\beta)$ is distributed symmetrically about θ if either of the following two conditions hold: (a) the distribution H is symmetric about some point $(\mu_1, \mu_2)'$. (b) the two fragmentary sample sizes are equal, i.e., s=t. Thus under the stated conditions, the estimator $\hat{\theta}$ is unbiased. Now, let us consider the asymptotic performance of $\hat{\theta}$.

<u>Theorem 1</u>. Let the distribution functions of X_1 and $Y_1-\theta-X_1$ be denoted by F and G. The corresponding density functions f and g of F and G are assumed to

satisfy the mild conditions $\int f^2(u)du<\infty$ and $\int g^2(u)du<\infty$, respectively. Also, let $N=2n+s+t$ and $\lambda_1,\lambda_2,\lambda_3$ be nonegative numbers such that $\lambda_1+\lambda_3>0$, $\lambda_2+\lambda_3>0$ and $\lambda_1+\lambda_2+2\lambda_3=1$. Then, as $N\to\infty$, $s/N\to\lambda_1$, $t/N\to\lambda_2$, $n/N\to\lambda_3$, the distribution of $N^{\frac{1}{2}}(\hat\theta(\alpha,\beta)-\theta)$ converges to a normal distribution with mean 0 and variance $\sigma^2(\hat\theta)$ =

$$[\alpha^2\{1/12+\lambda_3(1/2-\int F(u)F(v)dH(u,v))\}+\beta^2\lambda_3^3/\{12(\lambda_3+\lambda_1)(\lambda_3+\lambda_1)\}+$$

$$\alpha\beta\lambda_3^2\{1/2-2\int F(u)G(v-u)dH(u,v)\}/\{(\lambda_3+\lambda_1)(\lambda_3+\lambda_2)\}]/[\alpha(\lambda_3+\lambda_1)^{\frac{1}{2}}(\lambda_3+\lambda_2)^{\frac{1}{2}}$$

$$\int f^2(u)du+\beta\lambda_3^2\int g^2(u)du/\{(\lambda_3+\lambda_1)(\lambda_3+\lambda_2)\}].$$

The above theorem also shows that $\hat\theta(\alpha,\beta)$ is consistent. We note that the computation of $\hat\theta$ requires finding the median of the $\alpha(n+s)(n+t)+\frac{1}{2}\beta n(n+1)$ differences and becomes rather tedious for a large set of data. Fortunately, there are several shortcut methods of obtaining $\hat\theta$ (Lehmann, 1975).

3. THE EFFICIENCY OF $\hat\theta$

The estimator $\hat\theta$ is compared with the estimator $\hat\theta_1$ (Lin and Stivers, 1974, (2.4)), a regression estimator $\hat\theta_2$ (Gupta and Rohatgi, 1979, (9)) and a naive estimator $\hat\theta_3=\bar Y-\bar X$, where $\bar Y=\sum_{i=1}^{n+t} Y_i/(n+t)$ and $\bar X=\sum_{i=1}^{n+s} X_i/(n+s)$. All the estimators mentioned above are consistent and asymptotically normally distributed.

Let $\hat\theta$ and $\hat\theta'$ be asymptotically unbiased estimators for a parameter θ in the sense that both $N^{\frac{1}{2}}(\hat\theta-\theta)$ and $N^{\frac{1}{2}}(\hat\theta'-\theta)$ have asymptotic distributions with zero means. The asymptotic relative efficiency of $\hat\theta$ with respect to $\hat\theta'$, denoted by $ARE(\hat\theta,\hat\theta')$, is defined by

$$\text{ARE}\ (\hat{\theta}, \hat{\theta}') = \frac{\sigma^2(\hat{\theta}')}{\sigma^2(\hat{\theta})},$$

where $\sigma^2(\hat{\theta}) = \lim_{N \to \infty} \text{var}\ (N^{\frac{1}{2}}\hat{\theta})$ and $\sigma^2(\hat{\theta}') = \lim_{N \to \infty} \text{var}\ (N^{\frac{1}{2}}\hat{\theta}')$ (c.f. Randles and Wolfe (1979), p. 227).

Two bivariate distributions were considered for H in the comparison study:

(a) A bivariate normal distribution with unit variance and correlation coefficient ρ.

(b) A bivariate exponential distribution (Gumbel (1960)). The joint distribution function is

$$H(u,v) = F(u)F(v)\ [1+\tau\{1-F(u)\}\{1-F(v)\}],$$

$$\text{where } F(u)=1-e^{-u},\ -1\leq\tau\leq 1 \text{ and } \rho=\tau/4.$$

The advantage of using Gumbel's bivariate exponential distribution for our comparison study is that the asymptotic variance $\sigma^2(\hat{\theta})$ of $\hat{\theta}$ in Theorem 1 has a closed form. For each family of bivariate distributions mentioned above and a group of selected values of ρ and τ, the asymptotic relative efficiencies $e_i(\alpha,\beta)=\text{ARE}(\hat{\theta}(\alpha,\beta),\hat{\theta}_i)$, i=1,2,3, were computed. Table 1 provides the computational results for (A): λ_1=0.2, λ_2=.2, λ_3=0.3 and (B): λ_1=0.05, λ_2=.35, λ_3=.3 under the normal distribution. The estimator $\hat{\theta}$ (1,0) performs as well as $\hat{\theta}_i$(i=1,2,3) except for large values of ρ. The estimator $\hat{\theta}$(1,1) is more efficient than $\hat{\theta}$ (1,0) for positive ρ. This is not surprising because $\hat{\theta}$(1,0) ignores the pairing information. However, if the ratio β/α is too large, the efficiency of the estimator $\hat{\theta}(\alpha,\beta)$ becomes rather low except for extremely large ρ.

Results similar to Table 1 were also found for other combinations of λ_i, i=1,2,3. The choice of α=1 and β=1 generally yields a good estimator which can be applied for a wide range of values of ρ without requiring the user to specify or estimate the correlation coefficient ρ.

The estimator $\hat{\theta}$ does much better than $\hat{\theta}_i$, i=1,2,3, for heavy tail distributions as is to be expected because $\hat{\theta}_i$(i=1,2,3) is a linear combination of fragmentary sample means (See Table 2). In particular, if H is a bivariate Cauchy distribution (Johnson and Kotz, 1976, p. 295) which is not shown in our tables, the ARE $(\hat{\theta},\hat{\theta}_i)=\infty$, i=1,2,3.

4. REMARKS

An interesting feature of $\hat{\theta}(1,0)$ is that it can be used even when the pairing of X_i with Y_i, i= 1,...,n, cannot be identified (c.f. Hollander, Pledger, and Lin (1974)). For example, a statewide readiness test was given at the beginning of the 1979-80 school year to every incoming first grade public school student of South Carolina. The purpose of this test was to distinguish those students who were ready for the formal first grade curriculum from those who were not ready. A pilot testing was conducted to obtain the cutoff score using a random sample of South Carolina's kindergarten students at the end of the 1978-79 school year. Educators have constantly demonstrated that in the very early years of schooling a vast amount of a student's achievement is caused by maturation and not necessarily by instruction. This coupled with the fact that these two tests were conducted approximately four months apart establishes a concern as to how much the cutoff score previously determined by the pilot test should be moved upward. So the problem becomes estimating "maturational growth" occurred during the summer months. However, many of the students in the pilot test cannot be identified at the data analysis time due to various human factors. Therefore, all the parametric and regression estimation procedures mentioned before are not valid. A similar example was also cited by Hollander, Pledger and Lin (1974).

Theorem 1 can be used to construct a test of $\theta=\theta_0$, for any θ_0. By looking at the acceptance regions of such tests, one could—in theory—construct a confidence interval of θ based on the fragmentary samples.

5. APPENDIX

Proof of small sample properties of $\hat{\theta}$. The proof that the distribution of the difference $\hat{\theta}(\alpha,\beta)-\theta$ is free of θ is straightforward. The proof of the symmetry of the distribution of $\hat{\theta}(\alpha,\beta)$ about θ is similar to that given by Lehmann, 1975, Theorem 3, p. 86.

Two lemmas are needed to prove Theorem 1. First, let us define a scoring function ϕ for comparing two observations X_i and Y_j by

$$\phi(X_i,Y_j)=\begin{cases}1, & X_i<Y_j,\\ 0, & \text{otherwise,}\end{cases}$$

and

$$W_{X,Y}=\alpha\sum_{i=1}^{n+s}\sum_{j=1}^{n+t}\phi(X_i,Y_j)+\beta\sum_{i=1}^{n}\sum_{i'=1}^{n}\phi(X_i+X_i',\ Y_i+Y_i').$$

Lemma 1. For any real number c and integer i between 1 and $\alpha(n+s)(n+t)+\frac{1}{2}\beta n(n+1)$, the ith ordered difference $D_{(i)}\leq c$ if and only if $W_{X,Y-c}\leq\alpha(n+s)(n+t)+\frac{1}{2}\beta n(n+1)-i$.

Proof. C.f. Theorem 4 of Chapter 2, p. 87, of Lehmann, 1975.

Lemma 2. For $\theta=0$ and a positive real c, the distribution of

$$((n+s)(n+t)N)^{-\frac{1}{2}}\{W_{X,Y-c/\sqrt{N}}-\alpha(n+s)(n+t)p_1-\beta n(n+1)p_2/2\}$$

converges to a normal distribution with mean 0 and variance

$\eta^2=\alpha^2[1/12+\lambda_3\{1/2-2\int F(u)F(v)dH(u,v)\}]+\beta\ \lambda_3/\{12(\lambda_3+\lambda_1)(\lambda_3+\lambda_2)\}\ +$

$\alpha\beta\lambda_3^2\{1/2-2\int F(u)G(v-u)dH(u,v)\}/\{(\lambda_3+\lambda_1)(\lambda_3+\lambda_2)\},$

as $N\to\infty$, $s/N\to\lambda_1$, $t/N\to\lambda_2$, $n/N\to\lambda_3$, where $p_1=\int\bar{F}(u+c/\sqrt{N})dF(u)$, $\bar{F}(.)=1-F(.)$ and $p_2=\int G(u-2c/\sqrt{N})dG(u)$.

Proof. For convenience, let us define a sequence of fragmentary samples $(X_{1,m}, Y_{1,m})',\dots,(X_{n_m,m}, Y_{n_m,m})'$; $X_{n_m+1,m},\dots,X_{n_m+s_m,m}$; $Y_{n_m+1,m},\dots,$ $Y_{n_m+t_m\ m}$, where $(X_{i,m}, Y_{i,m}+c/\sqrt{N_m})'$ has distribution function H $(i=1,\dots,n_m,$ $N_m=2n_m+s_m+t_m)$, $X_{i,m}$ and $Y_{j,m}$ have distribution functions $F(x)$ and $F(y+c/\sqrt{N_m})$, $i=1,\dots,n_m+$, $j=1,\dots,n_m+t_m$, respectively, and $s_m/N_m\to\lambda_1$, $t_m/N_m\to\lambda_2$, $n_m/N_m\to\lambda_3$, as $m\to\infty$. Furthermore, we assume that $(X_{i,m},Y_{i,m})'$, $X_{n_m+j,m}$, and $Y_{n_m+k,m}$ are mutually independent for $i=1,\dots,n_m$, $j=1,\dots,s_m$, $k=1,\dots,t_m$ and $(X_{1,m},Y_{1,m})'=(U,V-c/\sqrt{N_m})'$, $m\geq 1$, where $(U,V)'$ has distribution function H.

Now, let $p_{1,m}=\int\bar{F}(u+c/\sqrt{N_m})dF(u)$ $\quad p_{2,m}=\int G(u-2c/\sqrt{N_m})dG(u)$,

$$W_m^*=\{(n_m+s_m)(n_m+t_m)\ N_m\}^{-\frac{1}{2}}\Big\{\alpha\sum_{i=1}^{N_m+s_m}\sum_{j=1}^{N_m+t_m}[\phi(X_{i,m},Y_{j,m})-p_{1,m}]\ +$$

$$\beta\sum_{i=1}^{n}\sum_{i'=1}^{n}[\phi(X_{i,m}+X_{i',m},\ Y_{i,m}+Y_{i',m}]-p_{2,m}\Big\}\ \text{and}$$

$$T_m^*=\alpha\left\{\frac{n_m+t_m}{(n_m+t_m)N_m}\right\}^{\frac{1}{2}}\sum_{i=1}^{n_m+s_m}\left\{\bar{F}(X_{i,m}+c/\sqrt{N_m})-p_{1,m}\right\}+\ \alpha\left\{\frac{n_m+s_m}{(n_m+t_m)N_m}\right\}^{\frac{1}{2}}\sum_{j=1}^{n_m+t_m}\left\{F(Y_{j,m})-p_{1,m}\right\}+$$

$$\frac{\beta n_m}{\{(n_m+s_m)(n_m+t_m)N_m\}^{\frac{1}{2}}} \sum_{i=1}^{n_m} \left\{G(Y_{i,m} - X_{i,m} - c/\sqrt{N_m}) - p_{2,m}\right\}.$$

Then, it follows from the argument provided by Hollander, Pledger and Lin (1974, p. 179) that W_m^* and T_m^* have the same limiting distribution.

The asymptotic normality of T_m^* follows from a version of Berry-Esséen Theorem (Chung, 1968, Theorem 7.1.2., p. 185).

Proof of Theorem 1. For any real c, $P_\theta\ (N^{\frac{1}{2}}(\hat\theta(\alpha,\beta)-\theta)<c)=P_0(N^{\frac{1}{2}}\hat\theta(\alpha,\beta)<c)=P_0(\hat\theta(\alpha,\beta)<c/\sqrt{N})$. Without loss of generality, we can assume that $\theta=0$. Consider first the case $\alpha(n+s)(n+t)+\frac{1}{2}\beta n(n+1)=2\ell+1$. By Lemmas 1 and 2 the fact that $\int f^2(u)du<\infty$ and $\int g^2(u)du<\infty$ (Olshen, 1967, and Mehra and Sarangi, 1967),

$$P_0(D_{(\ell+1)}<c/\sqrt{N})=P_0(W_{X,Y-c/\sqrt{N}} \le \frac{\alpha(n+s)(n+t)+\beta n(n+1)/2+1}{2})$$

$$\doteq P_0\Bigg[\{(n+s)(n+t)N\}^{-\frac{1}{2}}\left\{W_{X,Y-c/\sqrt{N}}-\alpha(n+s)(n+t)p_1-\beta n(n+1)p_2/2\right\}\le$$

$$\left(\frac{(n+s)(n+t)}{N}\right)^{\frac{1}{2}}\left\{\alpha N^{\frac{1}{2}}(1/2-\int \bar F(u+c/\sqrt{N})dF(u))+\frac{\beta\lambda_3^2\, N^{\frac{1}{2}}}{2(\lambda_3+\lambda_1)(\lambda_3+\lambda_2)}(1/2-\int G(u-2c/\sqrt{N})dG(u)\right\}\Bigg]$$

$$\longrightarrow \Phi\left[c\left\{\alpha(\lambda_3+\lambda_1)^{\frac{1}{2}}(\lambda_3+\lambda_2)^{\frac{1}{2}}\int f^2(u)du + \beta\lambda_3{}^2\int g^2(u)du/((\lambda_3+\lambda_1)(\lambda_3+\lambda_2))^{\frac{1}{2}}\right\}/n\right], \quad \text{(A-1)}$$

as $N\to\infty$, $s/N\to\lambda_1$, $t/N\to\lambda_2$. $n/N\to\lambda_3$, where Φ is the distribution function of $N(0,1)$.

In the case $\alpha(n+s)(n+t)+\frac{1}{2}\beta n(n+1)=2\ell$, the probability $P_0(\hat\theta\le c/\sqrt{N})$ is bounded below and above by $P_0(D_{(\ell+1)}\le c/\sqrt{N})$ and $P_0(D_{(\ell)}<c/\sqrt{N})$. By the same argument, it can be shown that these two probabilities have the same limiting value (A-1).

TABLE 1. THE ASYMPTOTIC RELATIVE EFFICIENCY $e_j(\alpha,\beta)$ X 100 UNDER BIVARIATE NORMAL DISTRIBUTION

(A): $\lambda_1=0.2,\ \lambda_2=0.2,\ \lambda_3=0.3$
(B): $\lambda_1=0.05,\ \lambda_2=0.35,\ \lambda_3=0.3$

Correlation coefficient	$e_1(1,0)$		$e_2(1,0)$		$e_3(1,0)$		$e_1(1,1)$		$e_2(1,1)$		$e_3(1,1)$	
ρ	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
-0.8	89	84	89	84	96	96	82	77	82	77	89	88
-0.6	91	89	93	89	96	96	85	82	86	83	90	89
-0.4	94	92	96	94	96	96	88	87	91	88	91	90
-0.2	95	95	99	97	96	96	91	91	95	93	92	92
0.0	96	96	101	98	96	96	94	94	99	97	94	94
0.2	94	94	100	97	95	95	96	96	101	99	97	97
0.4	89	88	94	90	94	94	96	95	102	98	102	103
0.6	77	75	82	77	94	94	93	91	98	94	113	115
0.8	53	50	55	51	94	94	81	79	84	80	143	148

TABLE 2. THE ASYMPTOTIC RELATIVE EFFICIENCY $e_j(\alpha,\beta)$ X100 UNDER BIVARIATE EXPONENTIAL DISTRIBUTION

(A): $\lambda_1=0.2,\ \lambda_2=0.2,\ \lambda_3=0.3$
(B): $\lambda_1=0.05,\ \lambda_2=0.35,\ \lambda_3=0.3$

τ	$e_1(1,0)$		$e_2(1,0)$		$e_3(1,0)$		$e_1(1,1)$		$e_2(1,1)$		$e_3(1,1)$	
	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
-0.8	287	286	299	293	290	290	243	240	254	246	245	243
-0.6	291	290	304	297	292	292	248	245	259	252	249	247
-0.4	294	293	308	300	294	294	253	251	265	257	253	251
-0.2	297	297	312	305	297	297	258	256	271	263	258	256
0.0	300	300	316	309	300	300	263	262	277	269	263	262
0.2	303	303	320	312	303	303	269	267	284	275	269	267
0.4	306	306	329	315	307	307	274	273	290	281	275	274
0.6	308	308	327	317	310	310	280	279	297	287	282	281
0.8	311	310	330	319	314	314	286	284	304	293	289	289

The author wishes to thank the referees for helpful comments on the manuscript. He also wishes to thank Mr. R. Garcia-Quintana for providing an interesting example, Professor P.E. Lin for calling the author's attention to a paper by Gupta and Rohatgi (1979) and Ms. Julie Paolella for her excellent typing. This research is partially supported by the U.S. Air Force Office of Scientific Research under Contract No. F49620-79-C-0140.

REFERENCES

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. J. Am. Statist. Assoc. 52, 200-214.

Chung, K. L. (1968). A Course in Probability Theory. New York: Harcourt Brace Jovanovich, Inc.

Gumbel, E. J. (1960). Bivariate exponential distributions. J. Am. Statist. Assoc. 55, 698-707.

Gumbel, E. J. (1961). Bivariate logistic distributions. J. Am. Statist. Assoc. 56, 335-349.

Gupta, A. K. and Rohatgi, V. K. (1981). Inference on the difference of means of correlated variables from fragmentary samples. Sankhya, Series A, to appear.

Hollander, M., Pledger, G., and Lin, P. E. (1974). Robustness of the Wilcoxon test to a certain dependency between samples. Ann. of Statist. 2, 177-181.

Hocking, R. R. and Smith, W. B. (1968). Estimation of parameters in the multivariate normal distribution with missing observations. J. Am. Statist. Assoc. 63, 159-173.

Johnson, N. L. and Kotz, S. (1976). Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley & Sons, Inc.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day, Inc.

Lin, P. E. (1971). Estimation procedures for difference of means with missing data. J. Am. Statist. Assoc. 68, 634-636.

Lin, P. E. (1973). Procedures for testing the difference of means with incomplete data. J. Am. Statist. Assoc. 68, 699-703.

Lin, P. E. and Stivers, L. E. (1974). On difference of means with incomplete data. Biometrika 61, 325-334.

Mehra, K. L. and Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments. Ann. Math. Statist. 38, 90-107.

Mehta, J. S. and Gurland, J. (1969). Some properties and an application of a statistic arising in testing correlation. Ann. Math. Statist. 40, 1736-1745.

Olshen, R. A. (1967). Sign and Wilcoxon tests for linearity. Ann. Math. Statist. 38, 1759-1769.

Randels, R. H. and Wolfe, D. A. (1979). Introduction to the Theory of Nonparametric Statistics. New York: John Wiley & Sons, Inc.

SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 81-0125 / 2. GOVT ACCESSION NO. AD-A095 869	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATION OF LOCATION DIFFERENCE FOR FRAGMENTARY SAMPLES	5. TYPE OF REPORT & PERIOD COVERED Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) L. J. Wei	8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0140
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of South Carolina, Department of Mathematics, Columbia, S. C. 29208	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of ScientificResearch/NM Bolling AFB, Washington, DC 20332	12. REPORT DATE December 1980
	13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Asymptotic relative efficiency; bivariate exponential distribution; consistent Hodges-Lehmann estimator; median.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A class of simple and robust estimators of the difference between location parameters of correlated variables is proposed when some observations on either of the variables are missing. We show that these estimators are consistent, asmptotically normally distributed, and insensitive to outlying observations Asymptotic relative efficiency comparisons with other known estimators are made to show the advantage of the proposed estimators.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
FILMED

- 8